Testimony

Advancing Trustworthy Artificial Intelligence

Jason Matheny

CT-A2824-1

Testimony presented before the U.S. House Committee on Science, Space, and Technology on June 22, 2023

RAND CORPORATION

For more information on this publication, visit www.rand.org/t/CTA2824-1.

Testimonies

RAND testimonies record testimony presented or submitted by RAND associates to federal, state, or local legislative committees; government-appointed commissions and panels; and private review and oversight bodies.

Published by the RAND Corporation, Santa Monica, Calif.

Limited Print and Electronic Distribution Rights

www.rand.org

Advancing Trustworthy Artificial Intelligence

Testimony of Jason Matheny[1]
The RAND Corporation[2]

Before the Committee on Science, Space, and Technology
United States House of Representatives

June 22, 2023

Chairman Lucas, Ranking Member Lofgren, and Members of the Committee: Good morning, and thank you for the opportunity to testify today. I'm the president and CEO of the RAND Corporation, a nonprofit and nonpartisan research organization. Before RAND, I served in the White House National Security Council and Office of Science and Technology Policy, as a commissioner on the National Security Commission on Artificial Intelligence, as assistant director of national intelligence, and as director of the Intelligence Advanced Research Projects Activity, which develops advanced technologies for the U.S. intelligence community. For the past 75 years, RAND has conducted research in support of U.S. national security and domestic policy. We manage four study and analysis federally funded research and development centers (FFRDCs) for the government focused on national and homeland security. Today, I'll focus my comments on how the federal government can advance AI in a beneficial and trustworthy manner for all Americans.

Among a broad set of technologies, AI stands out for both its rate of progress and its scope of applications. AI holds the potential to broadly transform entire industries, including ones critical

[1] The opinions and conclusions expressed in this testimony are the author's alone and should not be interpreted as representing those of the RAND Corporation or any of the sponsors of its research.

[2] The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest. RAND's mission is enabled through its core values of quality and objectivity and its commitment to integrity and ethical behavior. RAND subjects its research publications to a robust and exacting quality-assurance process; avoids financial and other conflicts of interest through staff training, project screening, and a policy of mandatory disclosure; and pursues transparency through the open publication of research findings and recommendations, disclosure of the source of funding of published research, and policies to ensure intellectual independence. This testimony is not a research publication, but witnesses affiliated with RAND routinely draw on relevant research conducted in the organization.

to our future prosperity. The United States is currently the global leader in AI;[3] however, AI systems have security and safety vulnerabilities, and a major AI-related accident in the United States could dissolve our lead, much like nuclear accidents set back the acceptance of nuclear power in the United States.

The United States can make safety a differentiator for our AI industry, just as it was a differentiator for our early aviation, automotive, and pharmaceutical industries. Government involvement in safety standards and testing led to safer products, which in turn led to consumer trust and market leadership. Today, government involvement can build consumer trust in AI that strengthens the U.S. position as a market leader. This is one reason why many AI firms are calling for government oversight to ensure that AI systems are safe and secure: It's good for business.

I will highlight five actions that the federal government could take to advance trustworthy AI:

1. Invest in potential research moonshots for trustworthy AI, including (1) generalizable approaches to evaluate the safety and security of AI systems before they are deployed, (2) fundamentals of designing agents that will persistently follow a set of values in all situations, and (3) microelectronic controls embedded in AI chips to prevent the development of large models that lack safety and security safeguards.
2. Accelerate AI safety and security research and development through rapid, high return-on-investment techniques, such as prize competitions. Prizes pay only for results and remove the costly barrier of researchers writing applications, making them a cost-effective way to pursue ambitious research goals while opening the field to nontraditional performers, such as small businesses.
3. Ensure that U.S. AI efforts conduct risk assessments prior to model training, as well as safety evaluations and red team tests prior to model deployment.
4. Ensure that the National Institute of Standards and Technology (NIST) has the resources needed to continue applications of the NIST Risk Management Framework, and fully participate in key international standards relevant to AI, such as ISO SC-42.
5. To prevent intentional or accidental misuse of advanced AI systems, (1) require that companies report the development or distribution of large AI computing clusters, training runs, and trained models (e.g., >1,000 AI chips, $>10^{26}$ operations, and >100 billion parameters, respectively); (2) include in federal contracts with cloud-computing providers requirements that they employ "know your customer" screening for all customers before training large AI models; and (3) include in federal contracts with AI developers "know your customer" screening, as well as security requirements to prevent the theft of large AI models.

[3] Although there are many ways to measure this, the Stanford Global AI Vibrancy Tool has consistently ranked the United States at the top. See Stanford University, "Global AI Vibrancy Tool: Who's Leading the Global AI Race?" undated, https://aiindex.stanford.edu/vibrancy/.